IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: Shum et al.

FILED VIA EFS ON October 10, 2007

Application No. 09/338,176

Filed: June 22, 1999

Confirmation No. 1062

For: METHOD AND APPARATUS FOR RECOVERING A THREE-DIMENSIONAL SCENE FROM TWO-DIMENSIONAL IMAGES

Examiner: Allen C. Wong

Art Unit: 2621

Attorney Reference No. 3382-52053-01

MAIL STOP APPEAL BRIEF – PATENTS
COMMISSIONER FOR PATENTS
P.O. BOX 1450
ALEXANDRIA, VA 22313-1450

APPEAL BRIEF

Sir:

This brief is in furtherance of the Notice of Appeal filed June 13, 2007. The fee required under 37 CFR 1.17(c) was submitted on August 13, 2007 when the Brief was initially filed.

I. REAL PARTY IN INTEREST

The real party in interest is Microsoft Corporation, by an assignment from the inventors recorded at Reel 010184, Frame 0776.

II. RELATED APPEALS AND INTERFERENCES

Currently, there are no other pending appeals or interferences known to appellant, the appellant's legal representatives, or assignees, which will directly affect or be directly affected by or have a bearing on the pending appeal.

Previously, on January 30, 2004, Applicants filed an appeal brief appealing the rejection of claims 1-37 of the instant application. On April 5, 2005, the Board of Patent Appeals and Interferences entered Decision on Appeal for Appeal No. 2004-2251 reversing the Examiner. This decision is included as Appendix B.

III. STATUS OF CLAIMS

Claim 37 remains rejected under 35 U.S.C. 102(b) as being anticipated by USP 5,729,471 to *Jain* et al. ("*Jain*") and remain appealed in conjunction with which the subject Appeal Brief is being filed. Claims 1-2, 4-9, 11-16, and 18-36 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Jain* in view of USP 5,612,743 to Lee ("Lee") and remain appealed in conjunction with which the subject Appeal Brief is being filed

IV. STATUS OF AMENDMENTS

A qualifying amendment was filed on December 8, 2006 and was entered. An amendment after final rejection was filed on May 10, 2007, but the amendment was not entered. Thus, for the purpose of Appeal the claims will be presented as they appeared after the entry of the amendment filed on December 8, 2006.

V. SUMMARY OF CLAIMED SUBJECT MATTER

The claims relate to a method of recovering a three-dimensional scene from two-dimensional images. The method comprises:

providing a sequence of frames; (see, e.g., Fig. 2, at 52; Fig. 3 at 62-67; page6, line 11 to page7, line 3; page 7, line 26 to page 8, line 6) dividing the sequence of frames into frame segments (see, e.g., Fig. 2, at 54; page 7, lines 7-12; page 8, lines 7-14) wherein the frames in the sequence comprise feature points (see, e.g., page 7, lines 7-8) and wherein the sequence of frames is divided into frame segments based upon frames in each frame segment having at least a minimum number of feature points being tracked to at least one base frame in the frame segment (see, e.g., Fig. 4; page 9, line 5 to page 10, line 18);

performing three-dimensional reconstruction individually for each frame segment derived by dividing the sequence of frames; (see, e.g., Fig. 2, at 56, Fig. 3; page 6, lines 13-17) and combining the three-dimensional reconstructed segments together to recover a three-dimensional scene for the sequence of images (see, e.g., Fig. 2 at 58; page 7, lines 18-25; page 20, line 25 to page 22, line 5.)

According to one embodiment (claim 9), dividing the sequence of frames into segments further includes wherein a segment includes a plurality of frames and wherein dividing includes: identifying a base frame, (see, e.g., Fig. 4 at 90, page 9, lines 5-9) identifying feature points in the base frame; and determining the segments such that every frame in a segment has at least a predetermined percentage of feature points identified in the base frame (see, e.g., Fig. 4 at 92; page 9, lines 10-22);

for each segment, encoding the frames in the segment into at least two virtual frames that include a three-dimensional structure for the segment and an uncertainty associated with the segment and wherein encoding includes choosing at least two frames in the segment that are at least a threshold number of frames apart (see, e.g., Fig. 3 at 74-76; page 8, lines 15-21);

for each of the at least two chosen frames, projecting a plurality of three-dimensional points into a corresponding virtual frame (see, e.g., Fig. 8 at 176; page 8, lines 15-21; page 18, line 20 to page 20, line 26) ; and

for each of the at least two chosen frames, projecting an uncertainty into the corresponding virtual frame (see, e.g., Fig. 8 at 178; page 8, lines 15-21; page 18, line 20 to page 20, line 26).

According to another embodiment (claim 22), also included is:

(a) identifying at least a first base frame in a sequence of two-dimensional frames; (see, e.g., Fig. 4 at 90; page 9, lines 5-9)

(b) adding the at least first base frame to create a first segment of frames of the sequence; (see, e.g., Fig. 4 at 90; page 9, lines 5-9)

(c) selecting feature points in at least the first base frame in the first segment of frames in the sequence; (see, e.g., Fig. 4 at 92, Fig. 6; page 9, lines 10-15; page 11, line 9 to page 12, line 16)

(d) analyzing a next frame in the sequence to identify the selected feature points in the next frame; (see, e.g., Fig. 4 at 94; page 9, line 23 to page 10, line 3)

(e) determining a number of the selected feature points from the base frame that are also identified in the next frame; (see, e.g., Fig. 4 at 96; page 10, lines 3-19) and

(f) if the number of the selected feature points from the base frame that are also identified in the next frame is greater than or equal to a threshold number, adding the next frame to the first segment of frames of the sequence; (see, e.g., Fig. 4 at 98; page 10, lines 3-19).

According to another embodiment (claim 31) also included is: an improvement comprising dividing a long sequence of frames into segments and reducing the number of frames in each segment by representing the segments using between two and five representative frames per segment, (see, e.g., Fig. 7 at 142 and 150; page 12, line 26 to page 13, line 18) wherein the representative frames are used to recover the three-dimensional scene and remaining frames are discarded so that the three-dimensional scene is effectively compressed (see, e.g., Fig. 7 at 142 and 150; page 12, lines 26-29; claim 31 as initially filed), wherein dividing the long sequence into segments includes identifying a base frame and tracking feature points between frames in the sequence and the base frame and ending a segment whenever a frame does not contain a predetermined threshold of feature points that are contained in the base frame (see, e.g., Fig. 4, at 94-106; page 3, lines 21-27; page 9, line 23 to page 10, line 11.)

According to another embodiment (claim 36), also included is:
calculating a partial model for each segment, wherein the partial model includes the same number of frames as the segment said partial model represents and wherein the partial model includes three-dimensional coordinates and camera pose, the camera pose comprising rotation and translation, for features within the frames (see, e.g., page 8, line 15-21; page 13, lines 19-21);

extracting virtual key frames from each partial model, the virtual key frames having three-dimensional coordinates for the frames and an uncertainty associated with the frames (see, e.g., Fig. 3 at 74 & 76; page 8, lines 22-28; page 18, line 20 to page 20, line 25); and

bundle adjusting the virtual key frames to obtain a complete three-dimensional reconstruction of the two-dimensional frames (see, e.g., Fig. 2 at 58, page 7, lines 18-25; Fig. 8 at 182, page 20, line 25 to page 22, line 5).

Claim 37 has means plus function steps. The identification of such steps and structure, material, or acts described in the specification is set forth below:

means for capturing two-dimensional images; (see, e.g., Fig. 1; Fig. 3 at 62-67; page 2, lines 1-9; page 4 line 28 to page 6, line 25; page 7 line 26 to page 8, line 6)

means for dividing the sequence into segments (see, e.g., Fig. 1, Fig. 2 at 54, Fig. 3 at 68, Fig. 4; page 4 line 28 to page 6, line 25; page 3, lines 10-12; page 7, lines 7-12; page 8, lines 7-14; page 9, line 5 to page 13, line 18)

means for calculating a partial model for each segment that includes three-dimensional coordinates and camera pose for features within the frames of the segment, the three-dimensional coordinates and camera pose being derived from the frames of the segment; (see, e.g., Fig. 1; Fig. 3 at 70, 72, Fig. 7; page 3, line 27 to page 4, line 3; page 4 line 28 to page 6, line 25; page 8, line 22 to page 9, line 3; page 12, line 17 to page 18 line 19.)

means for extracting virtual key frames from each partial model; (see, e.g., Fig. 1; Fig. 3 at 74 & 76; page 4 line 28 to page 6, line 25; page 8, line 22 to page 9, line 3; page 12, line 17 to page 17 line 4; page 18, line 20 to page 20, line 6.)

means for bundle adjusting the virtual key frames to obtain a complete three-dimensional reconstruction of the two-dimensional frames. (see, e.g., Fig. 1, Fig. 2 at 58, Fig. 7 at 160, Fig. 8 at 182, page 4 line 28 to page 6, line 25; page 7, lines 18-25, page 17, line 5 to page 18, line 19; page 20, line 7 to page 22, line 5).

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

Two issues are presented for review. The first issue is whether claim 37 (as amended in the amendment filed on December 8, 2006) is patentable under 35 U.S.C. § 102(e) over U.S.

Patent 5,729,471 to Jain et al., (*Jain*). The second issue is whether claims 1-2, 4-9, 11-16, and 18-36 (as amended in the amendment filed on December 8, 2006) are patentable under 35 U.S.C. § 103(a) over *Jain* in view of U.S. Patent 5,612,743 to Lee (*Lee*).

VII. GROUPING OF CLAIMS

Independent claims 1, 9, 23, 31, 36 and 37 and any of their respective dependent claims each contain different limitations that further distinguish each from the prior art. However, to facilitate the Board's consideration of this appeal, Applicants group the claims for the purposes of this appeal as follows:

For the purposes of this appeal only, patentability of claims 2 and 4-8 stand or fall with the patentability of claim 1, patentability of claims 11-16 and 18-22 stand or fall with the patentability of claim 9, patentability of claims 24-30 stand or fall with the patentability of claim 23, patentability of claims 32-35 stand or fall with the patentability of claim 31, claim 36 is in a group by itself, and claim 37 is in a group by itself,

VIII. ARGUMENT

1. Rejection of Claim 37 under 35 U.S.C. § 102 (b)

Applicants respectfully request reversal of the Examiner's rejection of claim 37 under 35 U.S.C. § 102(b) as being anticipated by *Jain*. To anticipate a claim, the reference must teach or suggest each and every element of the claim. *See* MPEP § 2131. More particularly, "A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference." *Verdegaal Bros. v. Union Oil Co. of California,* 814 F.2d 628, 631, 2 USPQ2d 1051, 1053 (Fed. Cir. 1987). In this case, the *Jain* reference cited by the Examiner fails to either explicitly or implicitly teach or suggest each and every element of the rejected claim.

A. The cited reference, *Jain,* does not teach or suggest each and every element of independent claim 37

Claim 37 is directed to a method of generating a three-dimensional scene from a sequence of two-dimensional images. More particularly, claim 37 recites, in part:

> means for capturing two-dimensional images;
> means for dividing the sequence into segments;
> *means for calculating a partial model for each segment that includes three-dimensional coordinates and camera pose for features within the frames of the segment, the three-dimensional coordinates and camera pose being derived from the frames of the segment*;
> *means for extracting virtual key frames* from each partial model; and
> *means for bundle adjusting the virtual key frames* to obtain a complete three-dimensional reconstruction of the two-dimensional frames. (Emphasis added).

1. The Examiner picks and chooses from different embodiments in Jain.

The Examiner asserts that *means for dividing the sequence into segments* is found in a first embodiment, while the Examiner asserts that *means for calculating a partial model <u>for each segment</u>* is found in a second embodiment. Under 35 USC § 102, every limitation of a claim must identically appear in a prior art reference for it to anticipate the claim. The Examiner must not cobble together disparate elements from separate embodiments to make such a rejection. The reference "must clearly and unequivocally disclose the claimed [invention] or direct those skilled in the art to the [invention] without any need for picking, choosing, and combining various disclosures not directly related to each other by the teachings of the cited reference." [In re Arkley, 455 F.d 586, 587, 175 USPQ 524, 526 (CCPA 1972).]

Jain discloses two main embodiments. To teach or suggest segments, the Examiner relies on Jain at Fig. 8, (showing that a camera 3 has 319 frames) and Jain 23:58-24:4, which describes the first embodiment, a "rudimentary, prototype MPI video system" (Jain, 24:22) in which, ideally, scene analysis (extraction of 3-D imagery from a single 2-D frame) "should be applied to every video frame," but, due to the amount of computational effort involved in doing so, scene analysis is limited to only one every thirty frames—the key frame. The other 29 frames are ignored. [Jain, 23:58-67.] In his rejection, the Examiner states: "clearly, Jain discloses there are segments within a sequence of frames, otherwise, the ascertainment of key frames would not be possible without these segments, where each segment is formed from a sequence of 30 frames." [*See* Office Action mailed March 13, 2007, at page 13.]

Thus, the Examiner equates the segment of claim 37 with Jain's description of processing only a portion of video frames, due to computational limitations. [Jain 23:61-63.]

The Examiner relies on the "image to ground projection" of Fig. 12 to teach or suggest a *partial model*. [*See* Office Action mailed March 13, 2007 at page 13, stating: "fig. 12, note there are multiple "image to ground projection" sections that are used to calculate and project an image or a partial model for each segment, quoted above."]

Figure 12 describes a second, vastly different embodiment from the first embodiment which the Examiner relies on to describe "segments." As the Examiner has picked and chosen from two separate embodiments to make this § 102 rejection, Applicants respectfully submit that the Examiner has failed to make a *prima facie* case against claim 1. As such, Applicants respectfully submit that the rejection of claim 37 under 35 U.S.C. § 102(b) is improper and claim 37 in its present form should be patentable.

2. Jain does not teach or suggest a partial model.

Jain fails to teach or suggest the claim 37 language *means for calculating a partial model for each segment that includes three-dimensional coordinates and camera pose for features within the frames of the segment.*

The Examiner rejects the above claim language as follows: "fig. 12, note there are multiple "image to ground projection" sections that are used to calculate and project an image or a partial model for each segment of that includes three-dimensional occupancy estimation for which a 3D map of is generated in an attempt to form a dynamic model; col. 21, ln.63 to col. 22, ln. 7, Jain discloses the use of equations that includes three dimensional coordinates (x, y, z) that includes camera position or pose, camera angle and camera parameter to obtain a partial model or a "image to ground projection.")" [*See* Office Action mailed March 13, 2007 at pages 12 and 13.]

a. Jain does not teach or suggest a partial model for each segment.

As a separate reason for patentability, The Examiner's idea of segments from the first embodiment are not described or otherwise discussed with reference to this second embodiment.

Rather, the portion of the first embodiment with which the Examiner equates with segments are taught away from.

This second embodiment is described as "an omniscient multi-perspective perception system based on multiple stationary video cameras" [Jain, 25:46-49.] that attempts to "capture, organiz[e] and process[] real-world events in order that a system action –such as, for example, an immediate selection, or synthesis, of an important video image (e.g., a football fumble, or an interception)—may be predicated on this detection." [Jain 25:30-34.] Thus, this model would be certainly expected to use the "ideal" analysis (as defined in the first embodiment) of processing every frame "to get the most precise information" as, if the "segments" (as defined by the Examiner) from the first embodiment are used, only one frame in thirty will be processed, with the effect that the "important video image" may be among those 29 frames that are discarded, and therefore missed by the system, completely negating its declared purpose—predicting important video images. [Jain 23:58-60.]

Further, as the second embodiment explicitly describes the significant computational resources used in the implementation, [Jain 30:24-55] the only reason for not processing every frame in the first embodiment, "significant human and computational effort" is not an issue, further strongly teaching away from processing only a portion of the available frames, and thus strongly teaching away from the Examiner's own definition of a segment, because if every frame is processed, then there is no segment generated.

As the second embodiment, in which the Examiner locates the "partial model" of claim 37, teaches away from what the Examiner defines as a "segment," Jain does not teach or suggest *calculating a partial model <u>for each segment</u>* as there is no segment (as defined by the Examiner) for which the partial model could be calculated. Thus, as Jain does not teach or suggest, at least the claim language "a partial model for each segment," the rejection of claim 37 is improper under U.S.C. § 102 (b).

b. The Image to Ground projection of Jain does not teach or suggest a partial model.

Moreover, the image to ground projection of Jain does not teach or suggest the "*partial model*" of claim 37. The Examiner equates the partial model with the "image to ground projection" in Figure 12 of Jain. [*See* Office Action mailed March 13, 2007at page 13, stating:

"fig. 12, note there are multiple "image to ground projection" sections that are used to calculate and project an image or a partial model for each segment."] The "image to ground projection" box shown in Fig. 12 is unmentioned within the specification of Jain.

Using an otherwise unexplained phrase – "image to ground projection" cannot be said to teach or suggest the different claim language "partial model." Further nothing about the phrase "image to ground projection" suggests a "partial model." Moreover, as the "image to ground projection" of Jain is not actually taught, in that there is no explanation about how to calculate the "image to ground projection" that Applicants can locate, Jain cannot be said to teach or suggest the claim language "*means for calculating* a partial model."

Additionally, calculating the partial model requires using "three-dimensional coordinates and camera pose for features *within the frames of the segment*." As Jain discloses processing a single frame out of thirty frames, which, if for argument's sake we assume is a segment that possesses a single undiscarded frame, Jain cannot teach or suggest the italicized claim language which requires, at a minimum, using *frames* of the segment. For this further reason, claim 37 is in condition for allowance.

3. Jain does not teach or suggest bundle adjusting.

Jain fails to teach or suggest the claim 37 language *means for bundle adjusting the virtual key frames* to obtain a complete three-dimensional reconstruction of the two-dimensional frames.

The Examiner rejects the above claim language as follows: "fig. 12, note the "3D visualization" section is the product of the adjusting of the virtual key frames to produce a complete three-dimensional reconstruction of the two dimensional frames obtained by video camera 1 to video camera N; also, col. 24, ln. 38-67, Jain discloses the key frames are used to obtain the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points from key frames, estimate or bundle adjustments were made to ascertain the best, possible three-dimensional reconstruction of the two dimensional frame data to yield the 3D visualization.")" [*See* Office Action mailed March 13, 2007 at page 14.]

a. Jain does not teach or suggest bundle adjusting the virtual key frames.

Claim 37 requires "bundle adjusting the virtual key frames." That is, the bundle adjustment is performed **on** the virtual key frames. The Examiner equates the "3D visualization" in Fig. 12 with bundle adjusting. [*See* Office Action mailed March 13, 2007 at page 14 which states: "fig. 12, note the "3D visualization" section is the product of the adjusting of the virtual key frames to produce a complete three-dimensional reconstruction of the two dimensional frames obtained by video camera 1 to video camera N."] In his rejection of the "virtual key frame" portion of claim 37, the Examiner previously equated the "Image to Ground Projection" of Fig. 12 with "virtual key frames." Therefore, Applicants assume the Examiner meant to associate "bundle adjusting the virtual key frames" with the action "performing 3D visualization on the Image to Ground Projections" presumably located in Fig. 12.

Even assuming (for argument only) that the "3D visualization" element is equivalent to "bundle adjusting," and the "Image to Ground Projection" is equivalent to a virtual key frame, as asserted by the Examiner, Fig. 12 of Jain still does not teach or suggest *bundle adjusting the virtual key frames*, as the Image to Ground Projections are not processed, or otherwise used by the "3D visualization. The specification of Jain does not otherwise mention either "3-D visualization" or "Image to Ground Projections." Therefore, the only information on the two is found within Fig. 12. In Fig. 12, the "3D visualization" element takes as input the "Dynamic Model," which, itself, takes as input, the "Global Object Tracker" and the "3D Occupancy Estimation." Further, the "Global Object Tracker" and the "3D Occupancy Estimation" take as input the "Image to Ground Projections." [Jain, Fig. 12.] Thus, even assuming that virtual key frames lurk within the "Image to Ground Projections" they would also have to remain unchanged through, at a minimum, the "Global Object Tracker" computations and the "Dynamic Model" computations to be operated on within the "3-D Visualization." Even though very little information is given as to the nature of the various operations shown in Fig. 12, it is self-evident that the data is transformed between boxes, such that whatever the image to ground projection is, the 3-D occupancy estimation, which takes the image to ground projection as input, is something else entirely. Thus, Fig.. 12 teaches away from even the most generous mapping of the claim language "bundle adjusting the virtual key frames."

Further, the slight information we have about the items in Fig. 12 of Jain teaches against the claim language. The "Global Object Tracker " is not mentioned in the Specification of Jain. Thus, we have no information about it, other than the name, which can hardly be said to teach or suggest "virtual key frames." Rather, the name suggests *objects* being *tracked* teaching away (or rather, is a completely different idea) from anything to do with frames, virtual or otherwise.

The Dynamic model is described as follows: "The dynamic model contains task specific information like two dimensional and three dimensional maps, dynamic objects, states of objects in the scene (e.g., a particular human is mobile, or the robot vehicle is immobile), etc." [Jain, 27:66-28:2.] The description here also sounds like image processing has already been done, that is, the frames were analyzed downstream somewhere, to determine the, e.g., "dynamic objects," "states of objects in the scene," etc, all of which suggest the processing of data over time, as otherwise the specific objects and the dynamic nature of those objects could not be ascertained. This strongly suggests that the data in the dynamic model has long been separated from the initial film frames, and as such, has nothing to do with any sort of process that relies on the existence of frames, let alone specific types of frames, such as "virtual key frames" and specific types of actions, such as "bundle adjusting" the "virtual key frames."

Nothing about the description of this, or any of these Fig. 12 items teaches, suggests, or even slightly hints at the notion of "bundle adjusting virtual key frames." For at least this further reason, claim 37 is in condition for allowance.

b. The 3D visualization of Jain does not teach or suggest bundle adjusting.

Moreover, the "3D visualization" of Jain does not teach or suggest bundle adjusting. The Examiner equates the bundle adjusting with the "3D visualization" in Figure 12 of Jain. [*See* Office Action mailed March 13, 2007 at page 14.] As previously mentioned, the "3D visualization" box shown in Fig. 12 is unmentioned within the specification of Jain.

Bundle adjusting is described, e.g., in the specification as follows:

> Bundle adjustment is a non-linear minimization process that is typically applied to all of the input frames and features of the input image stream. Essentially, bundle adjustment is a non-linear averaging of the features over the input frames to obtain the most accurate 3D structure and camera motion.). [Specification, page 2, lines 19-22.]

Claim language is given "the broadest reasonable interpretation consistent with the specification." *In re Bond*, 910 F.2d 831, 833, 15 USPQ2d 1566, 1567 (Fed. Cir. 1990) ("It is axiomatic that, in proceedings before the PTO, claims in an application are to be given their broadest reasonable interpretation consistent with the specification, ... and that claim language should be read in light of the specification as it would be interpreted by one of ordinary skill in the art.") Also see *In re Morris,* 127 F.3d 1048, 1054, 44 USPQ2d 1023, 1027 (Fed. Cir. 1997.) In re Bond, 910 F.2d 831, 833, 15 USPQ2d 1566, 1567 (Fed. Cir. 1990) ("As an initial matter, the PTO applies to the verbiage of the proposed claims the broadest reasonable meaning of the words in their ordinary usage as they would be understood by one of ordinary skill in the art, taking into account whatever enlightenment by way of definitions or otherwise that may be afforded by the written description contained in the applicant's specification.")

Therefore, "as claim language should be read in light of the specification as it would be interpreted by one of ordinary skill in the art" bundle adjustment would be understood by someone of ordinary skill in the art to comprise, at a minimum, a non-linear minimization process that is applied to frames, which "3D visualization" has nothing to do with. This level of analysis really isn't necessary, however, as an otherwise unexplained phrase – "3D visualization" cannot be said to teach or suggest the different claim language "bundle adjusting." Further, there is nothing about the phrase "3D visualization" that would lead one of skill in the art, or anyone else, to suspect that "bundle adjusting" was even remotely involved. Thus, Jain does not teach or suggest "bundle adjusting"

c. Estimating field markings in Jain does not teach or suggest bundle adjusting.

Further, Jain fails to teach or suggests "bundle adjusting the virtual key frames to obtain a complete three-dimensional reconstruction of the two-dimensional frames." The Examiner states "Jain discloses the key frames are used to obtain the best possible three dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points from key frames, estimates or bundle adjustments were made to ascertain the best, possible three-dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization)." [*See* Office Action mailed March 13, 2007 at page 14.] Thus, the Examiner is stating, we believe, that estimates disclosed in Jain at 24:51-67 teach or suggest "bundle adjusting" because the estimates help create the three-dimensional reconstruction in Jain. Applicants respectfully disagree.

As described above, "as claim language should be read in light of the specification as it would be interpreted by one of ordinary skill in the art" bundle adjustment would be understood by someone of ordinary skill in the art to comprise, at a minimum, a non-linear minimization process that is applied to frames. Jain's hand-estimating football field markings on film frames does not teach or suggest such a "bundle adjustment."

In the first embodiment in Jain, three cameras take film of a football game. To limit the amount of computation effort required, only one frame in thirty (the key frame) in the original film is analyzed. For each key frame, a person (ideally) locates field marks i.e., known real-world locations that will then be used to determine the location and orientation of the camera that took the original frame. [Jain, 23:63-64; 24:38-40, Fig. 9a] Some frames, however, didn't show the football field itself, and so the field mark locations on the field for those frames had to be "estimated", e.g., guessed. [See Fig. 9b for a slide which doesn't show the field, and thus cannot provide accurate field markings.] The estimated locations were then used as "known points" to determine camera status [Jain, 24:51-67, Fig. 9b.] That is, once estimated locations were marked on the frames, those locations were treated no differently than any of the other "known points" marked by hand to determine camera status.

To not belabor the point, bundle adjustment is "a non-linear minimization process" at a minimum, that has nothing to do with Jain's disclosure of hand-marking estimated field markings on frames. For this further reason, at least, claim 37 is in condition for allowance.

For, at least, the many reasons discussed above, claim 37 is in condition for allowance.

2. Rejection of Claims 1-2, 4-9, 11-16, and 18-36 under 35 U.S.C. § 103(a)

To establish a *prima facie* case of obviousness, three basic criteria must be met. First, for the claim under review as a whole, there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or to combine reference teachings. Second, there must be a reasonable expectation of success. Finally, the prior art reference (or references when combined) must teach or suggest all the claim limitations. 35 U.S.C. § 103(a); *In re Vaeck*, 947 F.2d 488, 20 USPQ2d 1438 (Fed. Cir. 1991).

A. The cited references, *Jain* and *Lee,* cannot be combined.

The combination of Jain and Lee proposed by the Examiner to reject claims 1-2, 4-9, 11-16, and 18-36 is improper. It is improper to combine Jain with Lee as the hand-drawn feature points of Jain cannot be substituted with the feature points in Lee, which are a computer-selected set of pixels that act in unison.

As to Jain, the Examiner states that "Jain does not specifically disclose the determining at least a minimum number of feature points being tracked." [*See* Office Action mailed March 13, 2007 at page 16.] The Applicants agree. The Examiner argues, however, that Lee teaches "the use of threshold values TH and comparison of threshold values of feature points between the current frame the reference frame to check if the threshold is exceeded, thus there is a minimum number of feature points that is determined. Therefore, it would have been obvious to one of ordinary skill in the art to combine the teachings of Jain and Lee, as a whole, for improving the encoding of video image data so as to accurately encode images via the selection of feature points according to the motion of objects in a financially robust manner." [*See* Office Action mailed March 13, 2007 at page 17.]

Applicants respectfully disagree. Jain cannot be modified as suggested by the Examiner.

If the proposed modification would render the prior art invention being modified unsatisfactory for its intended purpose, then there is no suggestion or motivation to make the proposed modification. In re Gordon, 733 F.2d 900, 221 USPQ 1125 (Fed. Cir. 1984). Also, if the proposed modification or combination of the prior art would change the principle of operation of the prior art invention being modified, then the teachings of the references are not sufficient to render the claims prima facie obvious. In re Ratti, 270 F.2d 810, 123 USPQ 349 (CCPA 1959). The proposed modification of Lee with Jain would render Jain unsatisfactory for its intended purpose, and would change the principle of operation of Jain, as the "feature points" described in Jain have nothing to do, other than language overlap with the "feature points" described in Lee.

The emphasis of Lee is to provide better techniques to compress transmitted video data by more efficiently encoding the motion vectors within the video data. Temporal redundancies occur between neighboring pixels in different frames. Motion estimation reduces temporal redundancy in successive video frames (interframes) by encoding "motion vectors," which predict how portions of the frame behave over several frames. Lee teaches encoding a single vector for a "feature point"—a series of pixels acting in unison—rather than encoding separate motion vectors for the individual pixels. The encoded motion vectors are then used to create the compressed video stream. [Lee, 1:15-2:56; FIG. 2.]

Jain, on the other hand has nothing to do with compressing video images. Rather, Jain describes manually processing a series of normal camera picture images to mark the locations of specific football players and specific football field markings. [Jain 22:11-15.] These hand-marked locations are called "feature points." Due to computational limitations, only one frame in every thirty, the key frames, have such hand-processing performed on it. [Jain 23:58-67.] Each key frame has "at least three field marks" chosen. [Jain 24:39-40.]

Thus, the two patents, Jain, and Lee, use similar language ("feature points") for describing entirely different concepts.

Specifically, in Lee, a "feature point" refers to "pixels which are capable of representing the motions of objects in the frame" [Lee 4:58-60] and are automatically extracted by comparing the difference between locations in successive frames and choosing regions whose pixels are different by a threshold amount. That is, locations that move together are determined to be features and are encoded similarly.

In Jain, a "feature point" is the location of a known player, or a known field mark, and is marked manually. A person determines, for example, where the player "Washington" is, and marks that location on a video frame. [Jain, 22:1-3; 22:19-22; 24:38-40.] These known feature points are then "used as known points in order to solve the three unknown parameters that determine camera status." [Jain, 24:38-43.]

The feature points of Lee, which are used to efficiently compress similar frames within a compressed video stream, have nothing to do, other than language overlap, with the feature points of Jain, which are hand-selected player and field locations. A mark on a frame indicating a player foot on a field marking has nothing to do with a set of pixels that represent the motion of objects within a frame. As the feature points in Jain mark known physical locations on a football field, they cannot be substituted for the feature points of Lee (the Examiner's suggested modification) as, at a minimum, the feature points of Lee do not mark a known physical location, and so cannot be "used as known points in order to solve the three unknown parameters that determine camera status." [Jain, 24:38-43.]

Applicants also respectfully note the following differences between the feature points of Jain and Lee, each of which would change the operation of Jain, and/or would render Jain unfit for its intended purpose. Each difference strongly teaches away from the combination.

A) Jain hand-marks a very small number of feature points. In Lee, the feature points are computer selected.

B) In Jain, each feature point is a *single* hand-marked location which indicates a known physical location on a football field. In Lee, *many* pixels that act in unison make up a feature point.

C) Jain makes it abundantly clear that computational limitations limit the number of frames and feature points that can be processed. Specifically, even though "scene analysis should be applied to every frame," only one of every 30 frames, for a total of only 36 frames were analyzed "due to the significant human and computational effort to do so." [*See* Jain 24:56-60; 23:56-63]. Each of these frames had but "three or more" feature points hand marked on them. [*See* Jain 24:38-40]. Even assuming, generously, that each frame had 6 feature points marked, that gives only a total of 216 feature points processed (36 frames * 6 feature points per frame). Conversely, Lee makes no such mention of computational limits. Rather, the systems and methods of Lee are astonishingly computationally intensive as the frames are processed on a pixel-by-pixel basis. [*See* Lee 4:6-7, 5:22-25.] Although the number of pixels for each frame is not mentioned in Lee, it is known by those of skill in the art that a standard 16mm frame of film, such as that found in Jain, has in excess of one million pixels. If we make the unrealistically conservative assumptions that there are only ten frames in Lee, and that there are only 200,000 pixels per frame, then the total number of pixel calculations are in excess of two million (10 frames * 200,000 pixels), three orders of magnitude greater than the calculations (216) in Jain. At any rate, the system of Lee, which performs analysis on a pixel-by-pixel basis could not more strongly teach away from the system of Jain which performs only a tiny number of calculations per frame, specifically due to computational limitations.

For at least these reasons, as Jain and Lee are not properly combined, claims 1-2, 4-9, 11-16, and 18-36 should be allowable.

B. The cited references, *Jain* and *Lee,* do not teach or suggest each and every element of independent claim 1.

Claim 1 is directed to a method of generating a three-dimensional scene from a sequence of two-dimensional images. More particularly, claim 1 recites:

> 1. A method of recovering a three-dimensional scene from two-dimensional images, the method comprising:
> providing a sequence of frames;
> dividing the sequence of frames into frame segments wherein the frames in the sequence comprise feature points and wherein the sequence of frames is divided into frame segments based upon frames in each frame segment having at least a minimum number of feature points being tracked to at least one base frame in the frame segment;
> performing three-dimensional reconstruction individually for each frame segment derived by dividing the sequence of frames; and
> combining the three-dimensional reconstructed segments together to recover a three-dimensional scene for the sequence of images. (Emphasis added).

Applicants respectfully request reversal of the Examiner's rejection of claim 1 under 35 U.S.C. § 103(a) as being obvious over *Jain* in view of *Lee*, for, e.g., the reasons stated below.

1. The Examiner has failed to provide a reference for the claim language "wherein the sequence of frames is divided into frame segments based upon frames in each frame segment having at least a minimum number of feature points *being tracked to at least one base frame in the frame segment.*"

A prima facie case of patentability requires, at a minimum, that all features of the claim be taught or suggested. *In re Vaeck*, 947 F.2d 488, 20 USPQ2d 1438 (Fed. Cir. 1991). The Examiner rejects the above claim language as follows: "Jain does not specifically disclose the determining at least a minimum number of feature points being tracked. However, Lee teaches the determining at least a minimum number of feature points being tracked." [*See* Office Action mailed March 13, 2007 at page 16, last paragraph.]

Applicants respectfully point out that the language of claim 1 recites "wherein the sequence of frames is divided into frame segments based upon frames in each frame segment having at least a minimum number of feature points *being tracked* ***to at least one base frame in the frame segment.*** " not merely "determining at least a minimum number of feature points being tracked" as stated by the Examiner. Moreover, the full rejection of the above claim language also does not mention any area of Jain that corresponds to the missing claim language, that applicants can locate. [*See* Office Action mailed March 13, 2007 at page 17, first paragraph.] Furthermore, the Applicant's representative has carefully read both Jain and Lee and can find no description in either that might teach or suggest "wherein the sequence of frames is divided into frame segments based upon frames in each frame segment having at least a minimum number of feature points ***being tracked to at least one base frame in the frame segment.***"

Since the Examiner has failed to provide a reference for at least one portion of claim 1, the rejection of claim 1 under 35 U.S.C. § 103(a) over Jain in view of Lee is improper and claim 1 in its present form should be patentable.

2. Lee does not teach or suggest "wherein the sequence of frames is *divided* into frame segments based upon frames in each frame segment *having at least a minimum number of feature points being tracked* to at least one base frame in the frame segment."

Lee does not teach or suggest **wherein the sequence of frames is *divided* into frame segments**

In his rejection of claim 1, the Examiner states: "Jain does not specifically disclose the determining at least a minimum number of feature points being tracked. However, Lee teaches the use of threshold values TH and comparison of threshold values of feature points between the current frame and the reference frame to check if the threshold is exceeded, thus, there is a minimum number of feature points that is determined." [Action of September 8, 2006, at pg. 16, 1st full paragraph.] The language of the claim requires a way to *divide* the frames into segments, as shown by the bolded features "the sequence of frames **is divided into frame segments** based upon frames in each frame segment having at least a minimum number of feature points being tracked to at least one base frame in the frame segment."

Lee does not have frame segments. The Examiner never teaches or suggests where in Lee there might be such segments. At most, the Examiner states that Lee teaches "determining at least a minimum number of feature points being tracked." [Action of September 8, 2006, at pg. 16, 1st full paragraph.] Even, if for argument's sake, we assume that Lee does teach as the Examiner states, that would still not teach or suggest a way to ***divide*** the frames into segments. As the Examiner does not disclose a reference which is even alleged to teach or suggest the above-bolded language, applicants respectfully state that the Examiner has not made a *prima facie* case, and, thus, for this additional reason, claim 1 is in condition for allowance.

Moreover, even if one were assume, for argument's sake, that Lee does take feature points being tracked, the combination still falls short. Lee does not teach segments, and as Jain only teaches a "segment" (for argument's sake) composed of 29 discarded frames and one processed frame, the combination produces a single frame, the non-discarded frame (not *frames*, as required by the claim language) with feature points being tracked.

In addition, Lee, as far as Applicants can tell, is silent on the number of feature points in each frame. Specifically, while Lee does discuss how to determine feature points, there is no minimum number of feature points mentioned for any frame that Applicants could locate. In fact, the word "minimum" does not appear within the disclosure of Lee.

Thus, Lee cannot teach or anticipate the claim language *"having at **least a minimum number** of feature points being tracked."*

As a separate reason for patentability, and as shown in section VIII.2.A, Jain and Lee cannot be properly combined.

For, at least, all the reasons listed above, claim 1 is in condition for allowance. Claims 2 and 4-8 depend on claim 1 and at least for the reasons set fort regarding claim 1, they also should be patentable over the cited references.

C. The cited references, *Jain* and *Lee,* do not teach or suggest each and every element of independent claim 9.

Claim 9 is directed to a method of generating a three-dimensional scene from a sequence of two-dimensional images. More particularly, claim 9 recites:

> A method of recovering a three-dimensional scene from two-dimensional images, the method comprising:
>
> identifying a sequence of two-dimensional frames that include two-dimensional images;
>
> dividing the sequence of frames into segments, wherein a segment includes a plurality of frames and wherein dividing includes: identifying a base frame, identifying feature points in the base frame; and determining the segments such that every frame in a segment has at least a predetermined percentage of feature points identified in the base frame;
>
> for each segment, encoding the frames in the segment into at least two virtual frames that include a three-dimensional structure for the segment and an uncertainty associated with the segment and wherein encoding includes choosing at least two frames in the segment that are at least a threshold number of frames apart;
>
> for each of the at least two chosen frames, projecting a plurality of three-dimensional points into a corresponding virtual frame; and
>
> for each of the at least two chosen frames, projecting an uncertainty into the corresponding virtual frame.

1. The Examiner has failed to provide a reference for the claim language *"wherein encoding includes choosing at least two frames in the segment that are at least a threshold number of frames apart"*

The Examiner has failed to provide a reference to either Jain or Lee which teaches or suggests the claim language, above, "that are at least a threshold number of frames apart…" The Examiner, on page 19, last paragraph from the bottom of the Office action of September 8, 2006, in his rejection of claim 9, quotes only a portion of the claim 9 language in his rejection--to wit: "for each segment, encoding the frames in the segment into at least two virtual frames that include a three-dimensional structure for the segment and an uncertainty associated with the segment and wherein encoding includes choosing at least two frames." However, Applicants respectfully point out that the language of claim 9 recites the additional feature "that are at least a threshold number of frames apart…" Further, the rejection of the claim language quoted by the examiner, found on page 20, first paragraph of the Office action of September 8, 2006, fails to remedy the situation, as it too fails to mention the feature "*that are at least a threshold number of frames apart….*"

A the Examiner has failed to provide a reference for at least the above feature of claim 9, the Examiner has failed to make a prima facie case against claim 9 and the rejection of claim 9 under 35 U.S.C. § 103(a) over Jain in view of Lee is improper.

2. Jain and Lee do not teach "for each segment, encoding the frames in the *segment into at least two virtual frames"*

Hand-selecting one frame out of each thirty-frame segment and then hand-drawing feature points on the frame does not teach or suggest "encoding the frames in the segment into at least two virtual frames." The Examiner relies on passages in Jain which refer to determining camera pose during a football game by choosing one frame out of every thirty produced, and then hand-marking three or more known field positions on the frame. The three known positions are then used to determine the orientation of the camera that originally shot the footage. [See Jain, 23:58-24:3; 24:38-67.]

The Examiner has previously stated that the portion of Jain that corresponds to the segments are the thirty frame sequence. [See the Office action of March 13, 2007, at page 19, second full paragraph.] Here, the Examiner is relying on the same teachings to teach this different claim language. This fails in a number of ways.

First, Jain does not teach or suggest "encoding the *frames* in the segment." Jain teaches using one frame out of thirty, and discarding the others. Therefore, even if we assume, for argument's sake, that the 30-frame sequence of Jain is a segment, there is only one frame within that sequence that is not discarded, and therefore available for encoding. This does not teach or suggest "encoding the *frames*" as there is at most, one frame to encode.

Next, Jain does not teach or suggest "encoding the frames in the segment *into at least two virtual frames.* " Jain at most, has one non-discarded frame per segment. Even if we accept, for argument's sake, that hand-drawing feature points on a frame is *encoding the frame*, as each segment consists of but a single frame, Jain cannot encode that frame into two different frames. Further, since the processing done on the key frame consists of marking points on the frame itself, this further teaches away from any action that would require somehow creating out of thin air another frame altogether. Moreover, there is no computation or process that Applicants can ascertain that would be benefited by extra frames.

Additionally, as Jain is very worried about computational limitations, (see, e.g., Jain 23:58-67) this teaches away from doing any more processing, such as would be required if extra frames were added. For this further reason, claim 9 is in condition for allowance.

Moreover, and using the same reasoning set forth with regard to claim 1 in section VIII.2.B.2, Lee (and Jain) fail to teach or anticipate "dividing the sequence of frames into segments, wherein a segment includes a plurality of frames and wherein dividing includes: identifying a base frame, identifying feature points in the base frame; and determining the segments such that every frame in a segment has at least a predetermined percentage of feature points identified in the base frame."

As a separate reason for patentability, and as shown in section VIII.2.A, Jain and Lee cannot be properly combined. For at least these reasons, Claim 9 in its present form, and its dependent claims 10-22, should be allowed.

D. The cited references, *Jain* and *Lee,* do not teach or suggest each and every element of independent claim 23.

Claim 23 is also directed to a method of recovering a three dimensional scene from a sequence of two-dimensional images. More particularly, claim 23 recites:

23. A method of recovering a three-dimensional scene from a sequence of two-dimensional frames, comprising:

(a) identifying at least a first base frame in a sequence of two-dimensional frames;

(b) adding the at least first base frame to create a first segment of frames of the sequence;

(c) selecting feature points in at least the first base frame in the first segment of frames in the sequence;

(d) analyzing a next frame in the sequence to identify the selected feature points in the next frame;

(e) determining a number of the selected feature points from the base frame that are also identified in the next frame; and

(f) if the number of the selected feature points from the base frame that are also identified in the next frame is greater than or equal to a threshold number, adding the next frame to the first segment of frames of the sequence.

2. Jain and Lee do not teach "*if the number of the selected feature points from the base frame that are also identified in the next frame is greater than or equal to a threshold number, adding the next frame to the first segment of frames of the sequence.*"

Jain fails to teach or suggest, at least, the claim 23 language "(f) *if the number of the selected feature points from the base frame that are also identified in the next frame is greater than or equal to a threshold number, adding the next frame to the first segment of frames of the sequence*."

In his rejection of claim 23, the Examiner states: "Jain does not specifically disclose the adding the second frame to the segment. However, Jain discloses the manual adjustment of the number of key frames, where the number is one key frame for every thirty frames, i.e. a segment (col. 23, Ln.64 to col.24, ln.3.). Therefore, since Jain teaches the manual adjustment of one key frame or representative frame for every thirty frames, it would have been obvious to one of ordinary skill in the art to manually change the number of key (representative) frames per segment from anywhere between two to five key or representative frames per segment if necessary for accurately enhancing the three-dimensional representation of the targeted scene." [*See* Office Action mailed March 13, 2007 at page 26, first whole paragraph.]

To make this rejection, the Examiner changes his definition of segment in mid-frame, with the two portions being incompatible.

Jain describes that "ideally the scene analysis process just described should be applied to every video frame in order to get the most precise information about (i) the location of the players and (ii) the events in the scene." [Jain, 23:58-61.] However, due to the "significant human and computational effort to do so" [Jain, 23:61-62] scene analysis is only applied to every 30th frame. [Jain, 23:64-67.] Thus, scene analysis is applied to a specific frame only. Due to the computational effort involved, Jain cannot perform the ideal scene analysis and process every frame, but instead can only process every 30th frame. The Examiner defines a segment as the one processed frame and the 29 unprocessed frames. (See, as quoted above, "However, Jain discloses the manual adjustment of the number of key frames, where the number is one key frame for every thirty frames, i.e. a segment.") The Examiner then states (as shown above) that "it would have been obvious to one of ordinary skill in the art to manually change the number of key (representative) frames per segment from anywhere between two to five key or representative frames per segment if necessary for accurately enhancing the three-dimensional representation of the targeted scene."

But, the Examiner originally defined a segment as a single frame with scene analysis applied together with the discarded frames around it. If, for the sake of argument, Jain allows more frames to have scene analysis applied due, perhaps, to greater computational power, then a segment will still consist of the single frame with scene analysis applied to it, but there will be fewer discarded frames between segments. This does not teach or suggest, but rather teaches against, *adding the next frame to the first segment of frames of the sequence,* as a segment would still consist of only of a single frame with a smaller number of frames discarded between it and the next key frame.. Further, there is no reason to add a next frame to the first segment of frames, as there is no use for another frame given within Jain, as each frame is treated separately, without reference to other frames.

Moreover, *selecting* a frame is different than *adjusting* the number of key frames. Nowhere does Jain suggest "adjusting" or changing the number of frames from which a key frame is chosen. Further, 30 frames is the standard NTSC frame rate for a second of film. [*See* Office Action mailed September 8, 2006 at page 3, line 22.] In Jain, a key frame is selected for each second (30 frames) of film. The number 30 was not chosen randomly, which leads away from adjusting or changing the number of frames (30) from which a key frame is chosen.

The Lee reference, either separately or in combination with Jain, also fails to teach or suggest the language of claim 23.

As a separate reason for patentability, and using the same reasoning set forth with regard to claim 1 in section VIII.2.B.2, Lee (and Jain) fail to teach or anticipate the claim 23 features "(e)determining a number of the selected feature points from the base frame that are also identified in the next frame; and (f) if the number of the selected feature points from the base frame that are also identified in the next frame is greater than or equal to a threshold number, adding the next frame to the first segment of frames of the sequence."

As a further reason for patentability, and as shown in section VIII.2.A, Jain and Lee cannot be properly combined. For at least all of the reasons mentioned above, claim 23 and its dependent claims 24-30 are in condition for allowance.

E. **The cited references, *Jain* and *Lee,* do not teach or suggest each and every element of independent claim 31.**

Claim 31 is directed to an improvement in the method of recovering a three-dimensional scene from a sequence of two-dimensional frames. More particularly, claim 31 recites as follows:

> 31. In a method of recovering a three-dimensional scene from a sequence of two-dimensional frames, an improvement comprising dividing a long sequence of frames into segments and reducing the number of frames in each segment by representing the segments using between two and five representative frames per segment, wherein the representative frames are used to recover the three-dimensional scene and remaining frames are discarded so that the three-dimensional scene is effectively compressed, wherein dividing the long sequence into segments includes identifying a base frame and tracking feature points between frames in the sequence and the base frame and ending a segment whenever a frame does not contain a predetermined threshold of feature points that are contained in the base frame.

The Examiner has failed to provide a prima facie case of anticipation for, e.g., the following features of claim 31:

"***reducing*** the number of frames in each segment ..."

"dividing a long sequence of frames into segments and *reducing* the number of frames in each segment by representing the **segments using between two and five representative frames per segment....**"

"**ending a segment** whenever a frame does not contain a predetermined threshold of feature points that are contained in the base frame....", and

"dividing the long sequence into segments includes identifying a base frame and tracking feature points between frames in the sequence and the base frame and ending a segment whenever a frame does not contain a predetermined threshold of feature points that are contained in the base frame." Each point will be taken in turn.

1. The Examiner has failed to provide a reference for the claim language "*reducing* the number of frames in each segment ... "

The Examiner has provided no reference for the claim language "***reducing*** the number of frames in each segment ..." The Examiner states: "Jain does not specifically disclose the reducing the number of frames in each segment by representing the segments using between two and five representative frames per segment. However, Jain discloses the manual adjustment of the number of key frames, where the number is one key frame for every thirty frames, i.e., a segment. [Office action of March 13, 2007, page 29, 2nd paragraph.] Applicants respectfully disagree. Without belaboring the point, as each segment (in Jain) comprises at most one frame, (the key frame) with the other 29 frames discarded, it is nonsensical to think of such segments being "reduced", as to do so would give a segment with no frames at all.

2. The Examiner has failed to provide a reference for the claim language "dividing a long sequence of frames into segments and *reducing* the number of frames in each segment by representing the *segments using between two and five representative frames per segment.* "

Jain and Lee, also, both fail to teach or suggest "dividing a long sequence of frames into segments and *reducing* the number of frames in each segment by representing the **segments using between two and five representative frames per segment**."

The Examiner states that "Jain does not specifically disclose the reducing the number of frames in each segment by representing the segments using between two and five representative frames per segment." [*See* Office Action mailed March 13, 2007 at page 29, lines 5-7.] Applicants agree. The Examiner then states: "However, Jain discloses the manual adjustment of the number of key frames, where the number is one key frame for every thirty frames, i.e., a segment. Therefore, since Jain teaches the manual adjustment of one key frame or representative frame for every thirty frames, it would have been obvious to one of ordinary skill in the art to manually change the number of key (representative) frames per segment from anywhere between two to five key or representative frames per segment if necessary for accurately enhancing the three-dimensional representation of the targeted scene." [*See* Office Action mailed March 13, 2007 at page 29, lines 7-14.] Applicants respectfully disagree.

The Examiner has not provided a reference which teaches "reducing the number of frames in each segment by representing the segments using between two and five representative frames per segment" as recited in claim 31. Applicants respectfully suggest that, at a minimum, an obviousness rejection should include a reference where the cited language is taught. "To establish *prima facie* obviousness of a claimed invention, all the claim limitations must be taught or suggested by the prior art. *In re Royka*, 490 F.2d 981, 180 USPQ 580 (CCPA 1974)." MPEP 2143.03. Since the cited references do not teach or suggest at least the cited portions of claim 31, Applicants respectfully suggest that this claim is in condition for allowance.

3. The Examiner has failed to provide a reference for the claim language "*ending a segment* whenever a frame does not contain a predetermined threshold of feature points that are contained in the base frame."

The Examiner states that "Jain does not disclose a predetermined threshold of feature points that are contained in the base frame" and that "Lee teaches the predetermined threshold of feature points that are contained in the base frame." [*See* Office Action mailed March 13, 2007 at page 29, lines 15-16.] Even, if for argument's sake, we assume that Lee does disclose predetermined feature points contained in a base frame, this neither teaches nor suggests the additional limitations "**ending a segment** whenever a frame does not contain a predetermined threshold of feature points that are contained in the base frame." The ending of segments is not mentioned in either Lee or Jain. As *ending a segment* under any circumstances is not taught, the additional limitations in bold "ending a segment **whenever a frame does not contain a predetermined threshold of feature points** that are contained in the base frame" are also not taught or suggested. As a 103 rejection requires, at a minimum, that all limitations be taught or suggested, claim 31 is in condition for allowance.

As a separate reason for patentability, and using the same reasoning set forth with regard to claim 1 in section VIII.2.B.2, Lee (and Jain) fail to teach or anticipate the claim 31 features "dividing the long sequence into segments includes identifying a base frame and tracking feature points between frames in the sequence and the base frame and ending a segment whenever a frame does not contain a predetermined threshold of feature points that are contained in the base frame."

As a separate reason for patentability, and as shown in section VIII.2.A, Jain and Lee cannot be properly combined.

For, at least, all the reasons given above, Claim 31 and its dependent claims 33-35 are in condition for allowance.

F. The cited references, *Jain* and *Lee,* do not teach or suggest each and every element of independent claim 36.

Claim 36 is directed to a computer-readable medium having computer-executable instructions for performing a method of recovering a three-dimensional scene from a sequence of two-dimensional frames. More particularly, claim 36 recites as follows:

> 36. A computer-readable medium having computer-executable instructions for performing a method comprising:
> providing a sequence of two-dimensional frames;
> dividing the sequence into segments;
> calculating a partial model for each segment, wherein the partial model includes the same number of frames as the segment said partial model represents and wherein the partial model includes three-dimensional coordinates and camera pose, the camera pose comprising rotation and translation, for features within the frames;
> extracting virtual key frames from each partial model, the virtual key frames having three-dimensional coordinates for the frames and an uncertainty associated with the frames; and
> bundle adjusting the virtual key frames to obtain a complete three-dimensional reconstruction of the two-dimensional frames.

Jain fails to teach or suggest the claim 36 language "calculating a partial model for each segment, **wherein the partial model includes the same number of frames as the segment said partial model represents** ... "

The Examiner has failed to provide a reference to either Jain or Lee which teaches or suggests the bolded portion of the claim language, above, "wherein the partial model includes the same number of frames as the segment said partial model represents ..." The Examiner, on page 31, last paragraph from the bottom of the Office action of March 13, 2007, in his rejection of claim 36, quotes only a portion of the claim 9 language in his rejection--to wit "calculating a partial mode for each segment that includes three-dimensional coordinates and camera pose for features within the frames, the camera pose comprising rotation and translation." However, Applicants respectfully point out that the language of claim 36 recites the additional feature "wherein the partial model includes the same number of frames as the segment said partial model represents ..." Further, the full rejection of the claim language quoted by the Examiner, found on page 32, first full paragraph of the Office action of March 13, 2007, fails to remedy the situation, as it too fails to mention the feature "wherein the partial model includes the same number of frames as the segment said partial model represents ...." As such, the Examiner has failed to make a *prima facie* case of obviousness, with claim 36 thus being in condition for allowance.

The reference to Jain also fails to teach or suggest many other aspects of Applicants' claim 36. For instance, Jain fails to teach or suggest calculating a partial model for each segment as discussed with reference to claim 37. Jain also fails to discuss extracting virtual key frames from each partial model, as discussed with reference to claim 37. Jain also fails to discuss "bundle adjusting the virtual key frames to obtain a complete three-dimensional reconstruction of the two-dimensional frames" also as discussed with reference to claim 37.

For at least these reasons, claim 36 is allowable.

As a separate reason for patentability, and using the same reasoning set forth with regard to claim 1 in section VIII.2.B.2, Lee (and Jain) fail to teach or anticipate the claim 36 features "dividing the long sequence into segments includes identifying a base frame and tracking feature points between frames in the sequence and the base frame and ending a segment whenever a frame does not contain a predetermined threshold of feature points that are contained in the base frame."

As a separate reason for patentability, and as shown in section VIII.2.A, Jain and Lee cannot be properly combined.

For, at least, all the reasons given above, Claim 36 is in condition for allowance.

IX. CONCLUSION

In light of the arguments presented above the rejection of claims 1, 2, 4-9, 11-16 and 18-37 should be reversed and all claims passed to issue.

Respectfully submitted,

KLARQUIST SPARKMAN, LLP

One World Trade Center, Suite 1600
121 S.W. Salmon Street
Portland, Oregon 97204
Telephone: (503) 595-5300
Facsimile: (503) 595-5301

By /Genie Lyons/
Genie Lyons
Registration No. 43,841

APPENDIX A

CLAIMS ON APPEAL

1. A method of recovering a three-dimensional scene from two-dimensional images, the method comprising:

providing a sequence of frames;

dividing the sequence of frames into frame segments wherein the frames in the sequence comprise feature points and wherein the sequence of frames is divided into frame segments based upon frames in each frame segment having at least a minimum number of feature points being tracked to at least one base frame in the frame segment;

performing three-dimensional reconstruction individually for each frame segment derived by dividing the sequence of frames; and

combining the three-dimensional reconstructed segments together to recover a three-dimensional scene for the sequence of images.

2. The method of claim 1 wherein performing includes creating at least two virtual key frames for each of the segments, wherein the virtual key frames are only a subset of the images in a segment but are a representation of all of the images in that segment.

3. (Canceled)

4. The method of claim 1 wherein performing further includes:

performing a two-frame structure-from-motion algorithm to create a plurality of local models for each segment; and

combining the plurality of local models by eliminating scale ambiguity.

5. The method of claim 4 further comprising:

bundle adjusting the combined local models to obtain a partial three-dimensional model for each segment;

extracting virtual key frames from the partial three-dimensional model, wherein the virtual key frames include three-dimensional coordinates for the images and an associated uncertainty; and

bundle adjusting all segments to obtain a complete three-dimensional model.

6. The method of claim 1 further including:

identifying feature points in the images;

estimating three-dimensional coordinates of the feature points; and

estimating a camera rotation and translation for a camera that captured the sequence of images.

7. The method of claim 1 wherein combining includes performing a non-linear minimization process across the different segments through bundle adjustment.

8. A computer-readable medium having computer-executable instructions for performing the method recited in claim 1.

9. A method of recovering a three-dimensional scene from two-dimensional images, the method comprising:

identifying a sequence of two-dimensional frames that include two-dimensional images;

dividing the sequence of frames into segments, wherein a segment includes a plurality of frames and wherein dividing includes: identifying a base frame, identifying feature points in the base frame; and determining the segments such that every frame in a segment has at least a predetermined percentage of feature points identified in the base frame;

for each segment, encoding the frames in the segment into at least two virtual frames that include a three-dimensional structure for the segment and an uncertainty associated with the segment and wherein encoding includes choosing at least two frames in the segment that are at least a threshold number of frames apart;

for each of the at least two chosen frames, projecting a plurality of three-dimensional points into a corresponding virtual frame; and

for each of the at least two chosen frames, projecting an uncertainty into the corresponding virtual frame.

10. (Canceled)

11. The method of claim 9 wherein the segments vary in length and wherein the length is associated with the number of frames in the segment.

12. The method of claim 9 further including:

identifying feature points in the sequence of two-dimensional frames;

estimating three-dimensional coordinates for the feature points; and

estimating camera rotation and translation for the feature points.

13. The method of claim 12 wherein estimating the three-dimensional coordinates includes applying a two-frame structure-from-motion algorithm to the sequence of two-dimensional frames.

14. The method of claim 9 further including:

dividing a segment into multiple frame pairs;

applying a two-frame structure-from-motion algorithm to the multiple frame pairs to create a plurality of local models; and

scaling the local models so that they are on a similar coordinate system.

15. The method of claim 14 wherein each of the multiple frame pairs includes a common base frame and one other frame in the segment.

16. The method of claim 15 further including interpolating frames between the multiple frame pairs.

17. (Canceled)

18. The method of claim 9 further including bundle adjusting the virtual frames from the segments to create a three-dimensional reconstruction.

19. The method of claim 9 further including identifying feature points in the frames by using motion estimation.

20. The method of claim 19 wherein the motion estimation includes:

creating a template block in a first frame including a feature point and neighboring pixels adjacent the feature point;

creating a search window used in a second frame; and

comparing an intensity difference between the search window and the template block to locate the feature point in the second frame.

21. The method of claim 9 wherein at most two virtual frames are used.

22. A computer-readable medium having computer-executable instructions for performing the method recited in claim 9.

23. A method of recovering a three-dimensional scene from a sequence of two-dimensional frames, comprising:

(a) identifying at least a first base frame in a sequence of two-dimensional frames;

(b) adding the at least first base frame to create a first segment of frames of the sequence;

(c) selecting feature points in at least the first base frame in the first segment of frames in the sequence;

(d) analyzing a next frame in the sequence to identify the selected feature points in the next frame;

(e) determining a number of the selected feature points from the base frame that are also identified in the next frame; and

(f) if the number of the selected feature points from the base frame that are also identified in the next frame is greater than or equal to a threshold number, adding the next frame to the first segment of frames of the sequence.

24. The method of claim 23 further including if the number of the selected feature points from the base frame that are also identified in the next frame is less than the threshold number, adding the next frame to a second segment of frames of the sequence and designating the next frame that falls below the threshold number as a second base frame in a second segment.

25. The method of claim 23 further including performing motion estimation to identify the feature points.

26. The method of claim 23 further including using corners as the feature points.

27. The method of claim 23 wherein the number of frames comprising a segment varies between segments.

28. The method of claim 23 further including creating two virtual key frames per segment.

29. The method of claim 28 further including bundle adjusting the virtual key frames of all the segments to obtain a three-dimensional reconstruction.

30. A computer-readable medium having computer-executable instructions for performing the method recited in claim 23.

31. In a method of recovering a three-dimensional scene from a sequence of two-dimensional frames, an improvement comprising dividing a long sequence of frames into segments and reducing the number of frames in each segment by representing the segments using between two and five representative frames per segment, wherein the representative frames are used to recover the three-dimensional scene and remaining frames are discarded so that the three-dimensional scene is effectively compressed, wherein dividing the long sequence into segments includes identifying a base frame and tracking feature points between frames in the sequence and the base frame and ending a segment whenever a frame does not contain a predetermined threshold of feature points that are contained in the base frame.

32. The method of claim 31 wherein each of the representative frames have an uncertainty associated therewith.

33. The method of claim 31 wherein the long sequence includes over 75 frames.

34. The method of claim 31 wherein dividing the long sequence into segments includes identifying a base frame and tracking feature points between frames in the sequence and the base frame and ending a segment whenever a frame does not contain a predetermined threshold of feature points that are contained in the base frame.

35. The method of claim 31 further including performing a two-frame structure-from-motion algorithm on each of the segments to create a partial model.

36. A computer-readable medium having computer-executable instructions for performing a method comprising:

providing a sequence of two-dimensional frames;

dividing the sequence into segments;

calculating a partial model for each segment, wherein the partial model includes the same number of frames as the segment said partial model represents and wherein the partial model includes three-dimensional coordinates and camera pose, the camera pose comprising rotation and translation, for features within the frames;

extracting virtual key frames from each partial model, the virtual key frames having three-dimensional coordinates for the frames and an uncertainty associated with the frames; and

bundle adjusting the virtual key frames to obtain a complete three-dimensional reconstruction of the two-dimensional frames.

37. An apparatus for recovering a three-dimensional scene from a sequence of two-dimensional frames by segmenting the frames, comprising:

means for capturing two-dimensional images;

means for dividing the sequence into segments;

means for calculating a partial model for each segment that includes three-dimensional coordinates and camera pose for features within the frames of the segment, the three-dimensional coordinates and camera pose being derived from the frames of the segment;

means for extracting virtual key frames from each partial model; and

means for bundle adjusting the virtual key frames to obtain a complete three-dimensional reconstruction of the two-dimensional frames.

APPENDIX B

RELATED APPEALS AND INTERFERENCES

Decision on Appeal for Appeal No. 2004-2251 reversing the Examiner on all claims.

MS/SYM/LCM
3382-52053-0

The opinion in support of the decision being entered today was *not* written for publication and is *not* binding precedent of the Board.

Paper No. 25

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Ex parte HEUNG-YEUNG SHUM, ZHENGYOU ZHANG, and QIFA KE

Appeal No. 2004-2251
Application No. 09/338,176

HEARD: April 5, 2005

Before MARTIN, JERRY SMITH, and BARRY, *Administrative Patent Judges.*

BARRY, *Administrative Patent Judge.*

DECISION ON APPEAL

A patent examiner rejected claims 1-37. The appellants appeal therefrom under 35 U.S.C. § 134(a). We reverse.

BACKGROUND

The invention at issue on appeal is aimed at reconstructing a three-dimensional ("3D") scene from a sequence of two-dimensional ("2D") images. (Spec. at 1.[1]) Structure-from-motion ("SFM") algorithms have been used to reconstruct such

[1] The appellants should number the lines of their specifications to facilitate specific citation thereto.

scenes. Aspects of SFM algorithms include "feature points," "baselines," and bundle adjustment. (*Id.* at 2.) A feature point is a point in an image that can be tracked well from one frame to another. Typically, corners of an object are considered good feature points. "The base line is associated with how a camera is moving in relation to an object depicted in an image." (*Id.*) "Bundle adjustment is a non-linear minimization process . . . typically applied to all of the input frames and features of the input image stream. Essentially, bundle adjustment is a non-linear averaging of the features over the input frames to obtain the most accurate 3D structure and camera motion." (*Id.*)

According to the appellants, "[t]here are . . . problems [associated] with conventional 3D reconstruction using SFM. For example, bundle adjustment of long sequence[s] of input frames may be computationally expensive if it involves processing the entire sequence of input frames and features at once." (Appeal Br. at 4.) Because "[t]he complexity of interleaving bundle adjustment for each iteration step may be measured as a function of the number of feature points and the number of frames being bundled," (*id.* at 5), they add, "bundle adjustment computed over a long sequence of input frames is time consuming and slows the entire 3D reconstruction." (*Id.*)

To overcome the shortcomings of conventional 3D reconstruction, the appellants' invention divides a long sequence of frames or images into smaller segments. "A 3D reconstruction is performed on each segment individually." (Spec. at 3.) "All the reconstructed segments are then combined . . . through an efficient bundle adjustment to complete the 3D reconstruction." (*Id.* at 29.) Because the complexity and, hence, the computational cost of 3D reconstruction and bundling are directly related to the number of frames being processed, the appellants assert that segmenting a longer sequence of frames reduces these costs. (Appeal Br. at 5.)

A further understanding of the invention can be achieved by reading the following claims.

> 1. A method of recovering a three-dimensional scene from two-dimensional images, the method comprising:
>
> providing a sequence of images;
>
> dividing the sequence of images into segments;
>
> performing three-dimensional reconstruction for each segment individually; and
>
> combining the three-dimensional reconstructed segments together to recover a three-dimensional scene for the sequence of images.
>
> 9. A method of recovering a three-dimensional scene from two-dimensional images, the method comprising:

> identifying a sequence of two-dimensional frames that include two-dimensional images;
>
> dividing the sequence of frames into segments, wherein a segment includes a plurality of frames;
>
> for each segment, encoding the frames in the segment into at least two virtual frames that include a three-dimensional structure for the segment and an uncertainty associated with the segment.

Claims 1-37 stand rejected under 35 U.S.C. § 102(e) as anticipated by U.S. Patent No. 6,046,745 ("Moriya").

OPINION

Rather than reiterate the positions of the examiner or the appellants *in toto*, we focus on the main point of contention therebetween. Observing that "Moriya's column 32, lines 40-46 disclose that it is possible to apply the Moriya's invention to any previously taken image or footage," (Examiner's Answer at 5), the examiner asserts, "the discussion of footage discloses the sequence of images that are obtained from the segment of a motion picture film that depicts a particular event since it would take a multitude or sequence of images to capture the whole essence, scene of a particular event." (*Id.*) The appellants argue, "*Moriya* refers to the term 'footage' only once in the entire patent and uses that term unmistakably in the context of the term's meaning as 'a single image' not 'a sequence of images' as claimed." (Reply Br. at 5.)

In addressing the point of contention, the Board conducts a two-step analysis. First, we construe claims at issue to determine their scope. Second, we determine whether the construed claims are anticipated.

1. Claim Construction

"Analysis begins with a key legal question — *what* is the invention *claimed*?" *Panduit Corp. v. Dennison Mfg. Co.*, 810 F.2d 1561, 1567, 1 USPQ2d 1593, 1597 (Fed. Cir. 1987). Here, independent claim 1 recites in pertinent part the following limitations: "[a] method of recovering a three-dimensional scene from two-dimensional images, the method comprising: providing a sequence of images. . . ." Independent claims 9, 23, 31, 36, and 37 recite similar limitations. Considering these limitations, the independent claims require recovering a 3D scene from a sequence of 2D images.

2. Anticipation Determination

"Having construed the claim limitations at issue, we now compare the claims to the prior art to determine if the prior art anticipates those claims." *In re Cruciferous Sprout Litig.*, 301 F.3d 1343, 1349, 64 USPQ2d 1202, 1206 (Fed. Cir. 2002). "A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference." *Verdegaal Bros., Inc. v. Union Oil Co.*, 814 F.2d 628, 631, 2 USPQ2d 1051, 1053 (Fed. Cir. 1987) (citing

Structural Rubber Prods. Co. v. Park Rubber Co., 749 F.2d 707, 715, 223 USPQ 1264, 1270 (Fed. Cir. 1984); *Connell v. Sears, Roebuck & Co.*, 722 F.2d 1542, 1548, 220 USPQ 193, 198 (Fed. Cir. 1983); *Kalman v. Kimberly-Clark Corp.*, 713 F.2d 760, 771, 218 USPQ 781, 789 (Fed. Cir. 1983)). "[A]bsence from the reference of any claimed element negates anticipation." *Kloster Speedsteel AB v. Crucible, Inc.*, 793 F.2d 1565, 1571, 230 USPQ 81, 84 (Fed. Cir. 1986).

Here, Moriya discloses "an image processing arrangement to determine camera parameters and make a three-dimensional shaped model of an object from a single frame picture image . . . in an interactive mode." Abs., ll. 1-4. Although the arrangement recovers a 3D scene, we are unpersuaded that it does so from a sequence of 2D images. To the contrary, the reference emphasizes that it operates on a single image. Specifically, "[i]t is important to note that within the present invention, determination of camera parameters, extraction of 3-D image data of objects and determination of 2-D CG image data are all **preferably and advantageously conducted using a single frame picture image (i.e., as opposed to having to use multiple differing frames** to determine camera parameters, etc.)." Col. 29, ll. 46-52 (emphasis added). Although the passage of Moriya cited by the examiner mentions "footage," it discloses that a single image can be drawn from such footage.

Specifically, "since the camera parameters can be determined directly from a **single image**, it is not necessary to set or record the camera parameters when **an image** is actually taken with a camera, and further, it is also possible to apply CG modelling to any previously taken **image** (e.g., vintage or historical footage) of which camera parameters are not known." Col. 32, ll. 40-46 (emphases added).

The absence of recovering a 3D scene from a sequence of 2D images negates anticipation. Therefore, we reverse the anticipation rejection of claim 1; of claims 2-8, which depend therefrom; of claim 9; of claims 10-22, which depend therefrom; of claim 23; of claims 24-30, which depend therefrom; of claim 31; of claims 32-35, which depend therefrom; and of claims 36 and 37.

CONCLUSION

In summary, the rejection of claims 1-37 under § 102(e) is reversed.

REVERSED

John C. Martin

JOHN C. MARTIN)
Administrative Patent Judge)
)
)
Jerry Smith)
) BOARD OF PATENT
JERRY SMITH) APPEALS
Administrative Patent Judge) AND
) INTERFERENCES
)
)
)
LANCE LEONARD BARRY)
Administrative Patent Judge)

APPENDIX C

EVIDENCE RELIED UPON BY THE APPELLANT IN THE APPEAL

None.